FINAL PERFORMANCE REPORT

Air Force Office of Scientific Research

BioInspired Concepts

Grant:
F49620-01-1-0564

Period covered (including extension):
July 1, 2001 – Feb. 28, 2005

Project:
"RELATIONSHIP BETWEEN NETWORK TOPOLOGY AND DYNAMICS:
CONCEPTS FROM GENOME-WIDE REGULATION OF CELL STATES"

Principal Investigator:
Sui Huang, M.D., Ph.D.
Vascular Biology Program, Department of Surgery
Karp Family Research Laboratories, Rm 11-212
Children's Hospital
300 Longwood Avenue
Boston, MA 02115
Phone: (617) 919-2393
Fax: (617) 730-0230
Sui.huang@tch.harvard.edu

Note:
This final performance report follows the structure given in the guideline "Performance Reports", BAA 2001-5

AFRL-SR-AR-TR-05-

0822

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	2005	FINAL July 2001 - Feb 2005

4. TITLE AND SUBTITLE
"RELATIONSHIP BETWEEN NETWORK TOPOLOGY AND DYNAMICS: CONCEPTS FROM GENOME-WIDE REGULATION OF CELL STATES"

5. FUNDING NUMBERS
G
AFOSR – BAA 2001-5
F49620-01-1-0564

6. AUTHOR(S)
Sui Huang, M.D., Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Children's Hospital Boston and Harvard
Medical School
Vascular Biology Program, Karp 11-212
300 Longwood Avenue
Boston, MA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This is the final performance report for a project funded through the BioInspired Concept program.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approve for Public Release: Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Despite accumulating data about the topology ("wiring diagram") of genome-wide gene regulatory networks little is known about the system behavior of such complex networks which produce the genome-wide molecular profiles that now can be measured with new genomic technologies. Using gene expression profiling to monitor multiple trajectories in genome-wise expression state space it was shown experimentally for the first time that a cell fate, in this case, neutrophil differentiation in a human cell line, represents a high-dimensional attractor state in gene expression state space - as predicted by theory. The use gene expression profiling for measuring trajectories necessitated the development of a computer program, GEDI to handle the new data structure. GEDI proved to be a powerful and versatile tool which allows intuitive and efficient inspection and analysis of large stacks of molecular profile data. Together with the insight in the constrained collective behavior of genes to attractors, this novel bioinformatics research tool will pave the way to applying molecular profiling methods for the high-throughput analysis of tissue response to unknown noxious agents. This in turn can be of great utility in the toxicological evaluation of new chemical and bioactive entities, including those relevant in chemical and biological warfare.

14. SUBJECT TERMS
Gene regulatory networks, attractors, gene expression profiles, GEDI

15. NUMBER OF PAGES
6 (incl. cover)

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

8-2-05

1. COVER SHEET - see standard form 298

2. OBJECTIVES

Overview of Specific Aims, their modification and redistribution of emphasis during the course of the project :

2.1. Original objective proposed in grant application of 2001:

Original Specific Aim 1 (THEORY): To analyze the dynamic behavior of random regulatory networks belonging to various topology classes (random, scale-free) in order to study the relationship between topology and dynamics

Original Specific Aim 2 (EXPERIMENTS): To use gene expression profiling techniques to demonstrate in mammalian cells features of network dynamics predicted by theoretical models, and to measure basic parameters

2.2. Newly defined Specific Aims

As mentioned in the annual progress reports the Aims were modified in the following way, reflecting (*1*) the rapid change in the research landscape at the interface between biology and physical sciences, notably, the emergence of Systems Biology and gene network analysis and (2) the necessity to develop a computational tool to achieve ***overall goal of the project***, which remained unchanged: *to promote understanding of the high-dimensional dynamics of complex gene regulatory networks.*

Specific Aim 1: **To analyze the topology of genome-wide molecular networks for structural bias and explain their evolution and functional significance**

Given the sudden and rapid rise of the field of analyzing the topology of genome-wide molecular networks, we have in an evasive manoeuver in response to rising academic competition, redefined the original Specific Aim 1 to make it more specific. Also, resources have been redistributed from Specific Aim 1 to the newly created Aim 3 (as first mentioned in annual report 2003). After submission of a manuscript (see section 6., Joy et al., 2005) which presents the results obtained under this Specific Aim, work on Specific Aim 1 was terminated in favor of the new Specific Aim 3.

Specific Aim 2: **To obtain through genome-wide gene expression profiling experimental evidence that the differentiated state in HL60 neutrophils is a high-dimensional stable attractor state in the genome-wide gene regulatory network**

This new Specific Aim 2 continues the original Specific Aim 2 and has become the main Specific Aim of the project driving other work. It has absorbed the lion share of efforts because it involves tedious experimentation and complex data analysis.

Specific Aim 3: **To port the GEDI program for integrated analysis of gene expression profile dynamics to JAVA and add features to enhance its functionality, and to evaluate expansion of its utility for clinical applications**

This new Specific Aim was defined after the decision to develop a platform-independent, user-friendly computational tool (GEDI) based on a pilot version that was created out of necessity for the analysis of data produced by the experimental work on Specific Aim 2. For the novel data structure (dynamic gene expression profiles of multiple parallel time courses) the existing analysis and representation tools were inappropriate. Pilot versions of the program turned out to be useful for a wide range applications for molecular profiling beyond the analysis of the dynamic data as those derived by the work in Specific Aim 2. In the second half of this 3-year project, work on this Specific Aim 3 had taken over the work on Aim1 in terms of priority and resources.

3. STATUS OF EFFORT

The Specific Aims of this project has been achieved with the publication of the results in three original research papers in 2003 and 2005 - corresponding to the three Specific Aims, as defined in section 2. For details, see section 4.

4. ACCOMPLISHMENT

Summary of achievements

1. ***From Specific Aim 1:*** Identification of a global network architecture feature in protein-interaction networks that was specific to biological systems and not found in simulated "scale-free networks". This insight will help understand the evolution and functional significance of the architecture of biological networks. The results have been published (**Joy et al., 2005**, see section 5.) and also led to the award of an ARO research grant in mathematics of complex systems.
2. ***From Specific Aim 2:*** New fundamental insight into system-level dynamics of complex molecular networks. The demonstration of the existence of high-dimensional attractors represents an important milestone for our understanding of mammalian cell regulation, such as cell fate determination in stem cells, and the origin of distinct, stable molecular profiles that may have clinical values. The results have been published in **Huang et al., 2005** (see section 5). For the concept of the approach, a patent application has been filed (see section 8).
3. ***From Specific Aim 3:*** Development of ***GEDI***, a user-friendly computer program for intuitive and efficient inspection and analysis of large volumes of static and dynamic biomolecular profiling data. A brief partial description of the underlying idea has been published (**Eichler et al., 2003**, see section 5). This program was awarded a U.S. **patent** and licensing negotiations with a biotech company is currently in progress (see sections 7. and 8.).

The following is a general overview underscoring the context, approach and significance.
For specific technical information, see publications listed ***in bold*** *in section 5 and patent disclosure.*

Despite the increasing knowledge about the architecture gene regulatory networks in the age of functional genomics, the integrated behavior of such networks that control cell behavior and thus, tissue homeostasis remains elusive. Yet such systems behavior of networks is important since they constrain the collective dynamics of gene expression and hence, explain the robust and recurrent, typical gene expression profiles associated with phenotypes or clinical conditions. Such genome-scale profiles of genetic activities can now be measured with DNA microarrays and appear to be useful for future diagnostics based on "signature profiles". The underlying hypothesis in this project is that the stable, observable gene expression profiles, such as those representing cell fates, are attractor states in a high-dimensional gene expression state space that represents the dynamics of the underlying genome-scale regulatory network. Attractors arise because of the dynamic constraints imposed by the gene regulatory interactions (i.e., the network) and epitomize the spontaneous order that can emerge in a complex, irregular network. Their existence have long been postulated based on computer simulations of random complex networks but never shown experimentally.

Experimental Approach. Using in vitro neutrophil differentiation as a case study, this project for the first time demonstrated the existence of high-dimensional attractors in the genomic regulatory networks and showed that they represent distinct cell fates. In the absence of knowledge of the network structure the presence of an attractor state can be inferred from the observation of contraction of a state space volume, which concretely translates into the measurable convergence of distinct gene expression profiles. To demonstrate such behavior, undifferentiated HL60 cells were

subjected to two pharmacologically perturbations, using dimethylsulfoxide and all-trans-retinoic acid which both stimulate neutrophil differentiation. These two biochemically distinct agents cause the cells to assume two distinct gene expression profiles, allowing the use of DNA microarray-based monitoring of gene expression profiles over 3000 genes to demonstrate the convergence of high-dimensional trajectories. This convergence is strongly indicative of a stable attractor state with respect to almost 3000 genes.

Computer program generated. The visualization of high-dimensional, multiple time course data necessitated the development of a new data visualization method, which led to the creation of GEDI (Gene Expression Dynamics Inspector), a program based on self-organizing maps (SOM) that represents the high-dimensional data as intuitive yet quantifyable color mosaics, each of them representing an individual sample/time point. Subsequent internal usability tests of GEDI among biologists and clinicians suggested that this program opens a new perspective to genome-wide analysis to non bioinformaticians. A U.S. patent has been issued for GEDI and it is being licensed to a biotech company but is freely available for academic and government use.

Significance for DoD. While the experiments showing that a cell fate may be an attractor of the underlying network represents an important step in our basic understanding of how genes govern system level cell behavior, such knowledge has also practical implications for the interpretation of molecular profiles of cellular responses in the near future. The notion of constrained dynamics with stable trajectories and attractor states may be particularly relevant in the classification and diagnosis of unknown, novel potentially noxious agents (e.g. in toxicology or biological warfare). The program GEDI unites machine learning with human brain's gestalt perception and will allow quick and efficient screening of large numbers of high-dimensional molecular profile data and the identification of patterns in the molecular response - thus complementing existing analysis that produces abstract numbers and lists as output. This holistic yet molecular and quantitative approach is particularly useful in explorative situations where no hypothesis exists before a measurement and opens the opportunity for non-bioinformaticians to inspect and make sense of molecular profiles. The latter may have great utility in military laboratories.

5. SCIENTIFIC PERSONNEL SUPPORTED

Scientists supported over the entire period or part of it. For % effort see annual reports.

Sui Huang, M.D., Ph.D., Principal Investigator,
Joy Maliackal Poulo, Ph.D., Postdoctoral fellow, physicist,
Gabriel Eichler, Graduate Student (currently at Boston University), bioinfomatician
Yuchun Guo, M.S., Computer scientist, software engineer

6. PEER-REVIEWED PUBLICATIONS

*Publications in **bold** represent those directly related to the three specific Aims and providing more detailed result descriptions.*

Huang S. Rational drug discovery: what can we learn from regulatory networks? Drug Discovery Today 2002; 7: S163-169.

Huang S. Regulation of cellular states in mammalian cells from a genome-wide view. In: Collado-Vides J, Hofestädt R, editors. Gene Regulation for Postgenomic Biology. Cambridge: MIT Press, 2002; p.181-220.

Eichler G., Huang S. and Ingber D.E. Gene Expression Dynamics Inspector (GEDI): A program for integrated analysis of expression profiles. Bioinformatics 2003; 19: 2321-2322.

Wilhelm T, Nasheuer HP, Huang S. Physical and functional modularity of the protein network in

yeast. Mol Cell Proteomics, 2003; 2: 292–298.
Huang S and Ingber DE. From Stem Cell to Functional Tissue Architecture: What are the signals and how are they processed? In: Stem Cell Handbook, Ed: Sell S. Humana Press, 2003.
Zhu H, Huang S, Dhar P. The next step in systems biology: simulating the temporospatial dynamics of molecular network. Bioessays 2003; 26:68-72.
Huang S. Back to the biology in Systems Biology: What can we learn from biomolecular networks? Brief. Functional Genomics Proteomics 2004; 2: 279-297
Zhu H, Huang S, Dhar P. Cellular automata with object-oriented features for parallel molecular network modeling. IEEE Transact Nanobioscience, in press
De Bivort B, Huang S, Bar-Yam Y. Dynamics of Cellular Level Function and Regulation Derived from Murine Expression Array Data., Proc Natl Acad Sci U S A, 2004; 101:17687-17692.
Maliackal JP., Brock A., Ingber DE and Huang, S. High "betweenness" proteins in the yeast protein interaction network. J. Biomed. Biotechnol. 2005; in press
Huang S. Multistability and Multicellularity: Cell Fates as High-dimensional Attractors of Gene Regulatory Networks. In: Computational Systems Biology, Eds: Kriete A, Eils R. Elsevier Acadmic Press, in press
Huang S, Eichler G, Bar-Yam Y, Ingber D. Cell fate as a high-dimensional attractor of a complex gene regulatory network. Phys. Rev Lett 2005; 94:128701
Huang S, Brangwynne CP, Parker KK, Ingber DE. Symmetry-breaking in mammalian cell cohort migration during tissue pattern formation: role of random-walk persistence. Cell Motil. Cytoskeleton 2005, 61:201-13

7. INTERACTIONS, TRANSITIONS

Negotiations for licensing with BG Medicine, Waltham, MA (http://www.bg-medicine.com/index.asp) are under way as of this moment.

8. NEW DISCOVERIES, INVENTIONS, OR PATENT DISCLOSURES

Patent pending: Huang S, Ingber DE, inventors. "Methods for Analyzing Dynamic Changes in Cellular Informatics and Uses Therefor " (09/985,963)

Patent issued: Eichler G, Huang S, Ingber DE. "A method and apparatus for displaying information" (10/435,660)

The GEDI program can be downloaded at:
http://www.childrenshospital.org/research/ingber/GEDI/gedihome.htm
and is free for academic and government use.

AFRL-SR-AR-TR-05-

0322

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	2005	FINAL July 2001 - Feb 2005

4. TITLE AND SUBTITLE
"RELATIONSHIP BETWEEN NETWORK TOPOLOGY AND DYNAMICS: CONCEPTS FROM GENOME-WIDE REGULATION OF CELL STATES"

5. FUNDING NUMBERS
G
AFOSR – BAA 2001-5
F49620-01-1-0564

6. AUTHOR(S)
Sui Huang, M.D., Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Children's Hospital Boston and Harvard
Medical School
Vascular Biology Program, Karp 11-212
300 Longwood Avenue
Boston, MA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This is the final performance report for a project funded through the BioInspired Concept program.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approve for Public Release: Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Despite accumulating data about the topology ("wiring diagram") of genome-wide gene regulatory networks little is known about the system behavior of such complex networks which produce the genome-wide molecular profiles that now can be measured with new genomic technologies. Using gene expression profiling to monitor multiple trajectories in genome-wise expression state space it was shown experimentally for the first time that a cell fate, in this case, neutrophil differentiation in a human cell line, represents a high-dimensional attractor state in gene expression state space - as predicted by theory. The use gene expression profiling for measuring trajectories necessitated the development of a computer program, GEDI to handle the new data structure. GEDI proved to be a powerful and versatile tool which allows intuitive and efficient inspection and analysis of large stacks of molecular profile data. Together with the insight in the constrained collective behavior of genes to attractors, this novel bioinformatics research tool will pave the way to applying molecular profiling methods for the high-throughput analysis of tissue response to unknown noxious agents. This in turn can be of great utility in the toxicological evaluation of new chemical and bioactive entities, including those relevant in chemical and biological warfare.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Gene regulatory networks, attractors, gene expression profiles, GEDI	6 (incl. cover)
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

8-2-05